Common inheritance of chromosome Ia associated with clonal expansion of *Toxoplasma gondii*

Asis Khan,[1,10] Ulrike Böhme,[2,10] Krystyna A. Kelly,[3,4,10] Ellen Adlem,[2] Karen Brooks,[2] Mark Simmonds,[2] Karen Mungall,[2] Michael A. Quail,[2] Claire Arrowsmith,[2] Tracey Chillingworth,[2] Carol Churcher,[2] David Harris,[2] Matthew Collins,[2] Nigel Fosker,[2] Audrey Fraser,[2] Zahra Hance,[2] Kay Jagels,[2] Sharon Moule,[2] Lee Murphy,[2] Susan O'Neil,[2] Marie-Adele Rajandream,[2] David Saunders,[2] Kathy Seeger,[2] Sally Whitehead,[2] Thomas Mayr,[3] Xuenan Xuan,[5] Junichi Watanabe,[6] Yutaka Suzuki,[7] Hiroyuki Wakaguri,[7] Sumio Sugano,[7] Chihiro Sugimoto,[5] Ian Paulsen,[8] Aaron J. Mackey,[9] David S. Roos,[9] Neil Hall,[2,8] Matthew Berriman,[2] Bart Barrell,[2] L. David Sibley,[1] and James W. Ajioka[3,11]

[1]Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, Missouri 63110, USA; [2]The Wellcome Trust Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, CB10 1SA, United Kingdom; [3]Department of Pathology, University of Cambridge, Cambridge, CB2 1QP, United Kingdom; [4]MRC Biostatistics Unit, Institute of Public Health, Cambridge, CB2 2SR, United Kingdom; [5]National Research Center for Protozoan Diseases, Obihiro University of Agriculture and Veterinary Medicine, Inada-cho, Obihiro, Hokkaido, 080-8555, Japan; [6]Department of Parasitology, Institute of Medical Science, The University of Tokyo, 4-6-1, Shirokanedai, Minato-ku, Tokyo, 108-8639, Japan; [7]Department of Medical Genomics, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8562, Japan; [8]The Institute for Genomic Research, Rockville, Maryland 20850, USA; [9]Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

Toxoplasma gondii is a globally distributed protozoan parasite that can infect virtually all warm-blooded animals and humans. Despite the existence of a sexual phase in the life cycle, *T. gondii* has an unusual population structure dominated by three clonal lineages that predominate in North America and Europe, (Types I, II, and III). These lineages were founded by common ancestors ~10,000 yr ago. The recent origin and widespread distribution of the clonal lineages is attributed to the circumvention of the sexual cycle by a new mode of transmission—asexual transmission between intermediate hosts. Asexual transmission appears to be multigenic and although the specific genes mediating this trait are unknown, it is predicted that all members of the clonal lineages should share the same alleles. Genetic mapping studies suggested that chromosome Ia was unusually monomorphic compared with the rest of the genome. To investigate this further, we sequenced chromosome Ia and chromosome Ib in the Type I strain, RH, and the Type II strain, ME49. Comparative genome analyses of the two chromosomal sequences revealed that the same copy of chromosome Ia was inherited in each lineage, whereas chromosome Ib maintained the same high frequency of between-strain polymorphism as the rest of the genome. Sampling of chromosome Ia sequence in seven additional representative strains from the three clonal lineages supports a monomorphic inheritance, which is unique within the genome. Taken together, our observations implicate a specific combination of alleles on chromosome Ia in the recent origin and widespread success of the clonal lineages of *T. gondii*.

Toxoplasma gondii is a globally distributed protozoan parasite that infects ~25% of the human population and a similar proportion of birds and mammals (Dubey and Beattie 1988; Hall et al. 2001). Acute infection during pregnancy can result in birth defects and the parasite may cause fatal encephalitis in immunocompromised individuals (Hall et al. 2001). However, the vast majority of infections go unnoticed or cause mild flu-like symptoms, resulting in an asymptomatic, chronic infection lasting the lifetime of the host. Importantly, the relatively quiescent tissue cyst form of the parasite in the chronic infection is readily transmitted between intermediate hosts by carnivorous or omnivorous feeding (Dubey and Beattie 1988). This is thought to be the driving force behind the recent and rapid expansion of *T. gondii* (Su et al. 2003).

By their very nature, parasitic organisms have evolved com-

plex life cycles that exploit their intimate association with their hosts. For apicomplexan species, including the causative agents of malaria (*Plasmodium* spp.) and avian coccidiosis (*Eimeria* spp.), the maintenance of a sexual cycle may be vital for generating the diversity necessary to adapt to varying environments by evading host immunity and optimizing transmission and virulence (Kyes et al. 2001; Mackinnon and Read 2004). The cost of this diversity comes at the expense of being reliant on a single definitive host and the more general problem of breaking up advantageous combinations of alleles (Maynard-Smith 1978). *T. gondii*, arguably the most successful species in the phylum, is a striking exception because, along with the sexual cycle, the parasite can transmit asexually by ingestion of tissue cysts by intermediate hosts (Dubey and Beattie 1988). This property, termed asexual oral transmission, underpins its highly clonal population structure with only three predominant lineages (Types I, II, and III) widespread in North America and Europe, where members within each lineage are virtually identical (Darde et al. 1992; Sibley and Boothroyd 1992; Howe and Sibley 1995). Equally remarkable, when comparing between lineages, there exist only two alleles at each locus. Together, these patterns indicate that the clonal lineages arose very recently from a common set of parental strains (referred to as Adam and Eve or A&E) and subsequently underwent only limited genetic exchange (Grigg et al. 2001a; Su et al. 2003). Thus, the need to define the mechanisms and common genetic elements that underlie asexual transmission and the ubiquitous distribution of *T. gondii* provides a compelling rationale for genetic and comparative genomic analyses.

Experimental genetic crosses between prototypic members of the three common lineages demonstrate a Mendelian pattern of inheritance (Su et al. 2002; Khan et al. 2005). There are 14 distinct linkage groups that comprise the ~65 Mb haploid genome. The recently compiled genome map relies on the presence of frequently occurring A&E polymorphisms for 229 well-dispersed genetic markers (Su et al. 2002; Khan et al. 2005); polymorphisms occur at a frequency of ~1 in 100 bp on all chromosomes with the exception of chromosome Ia (*chr*Ia). To our surprise, and unlike any other chromosome, genetic polymorphisms proved exceedingly difficult to identify on *chr*Ia. Moreover, the density of polymorphic markers elsewhere in the genome (spaced approximately every 300 kb) suggests that *chr*Ia is unique in its lack of polymorphism compared with the rest of the genome. This lack of polymorphism has been confirmed by a recent global analysis of single nucleotide polymorphisms (SNPs) identified by mapping ESTs against the whole-genome sequence of *T. gondii*. This analysis showed that the genome-wide polymorphism rate between the three lineages is ~0.65%, while it is only ~0.03% on *chr*Ia (Boyle et al. 2006). These findings prompted us to generate and analyze the finished genomic sequence for *chr*Ia and the similarly sized, but independent chromosome Ib (*chr*Ib) from the RH (Type I) strain. We subsequently compared nucleotide composition, repeats, and genes between the two chromosomes. Analyses of the levels of polymorphisms were used to address directly and accurately the oddity of apparent low polymorphism specific to *chr*Ia, by comparison with the ME49 clone B7 (Type II) strain. Sequencing of several regions at intervals across *chr*Ia for an additional seven representative strains, including Type III strains, was also used to explore the generality of the observation that *chr*Ia appears to be monomorphic.

Results

Comparison of *chr*Ia and *chr*Ib in the Type I strain RH reveals similar features

To provide a basis for initial comparative genomic analysis, we undertook complete sequencing of *chr*Ia and *chr*Ib genomic DNA from the RH strain of *T. gondii*. The genomic sequence for the RH strain reveals that the 1.92-Mb *chr*Ia and the 2.01-Mb *chr*Ib have an average G+C content of just over 52% with 211 protein-coding sequences (CDS) predicted for *chr*Ia and 252 for *chr*Ib (Table 1). Expression of 47% of the total predicted genes was confirmed by comparison with an extensive set of Expressed Sequence Tag (EST) data available at http://www.cbil.upenn.edu/apidots/downloads/ (Wan et al. 1996; Ajioka et al. 1998), of which 35% were identified completely. Comparative genome analyses revealed that both *chr*Ia and *chr*Ib have similar levels of coding and noncoding regions and contain conventional telomeres (Table 1; Supplemental material). Crucially, *chr*Ia does not appear to be unusual in structure or content compared with *chr*Ib, which we have used to be representative of the rest of the *T. gondii* genome.

Chromosomal comparison between RH and ME49 clone B7 reveal virtually identical *chr*Ia

Having completed the analysis of the sequence of the Type I RH strain, we were able to compare it with the recently completed whole-genome shotgun analysis of the Type II ME49 strain undertaken at TIGR (available at http://www.tigr.org and http://www.toxodb.org/ToxoDB.shmtl). Genomic sequence alignments between the Type I RH and Type II ME49 clone B7 for both chromosomes show that each chromosome pair has the same gene content and structural elements. However, there is a very dramatic difference in the levels of polymorphism between the two chromosomes (Fig. 1; Supplemental Table 2): *chr*Ib has 12,925 SNPs compared with only 332 SNPs on *chr*Ia (Table 1). An analysis of the SNPs in the intron sequences of a sample of genes from different chromosomal locations estimates that about seven SNPs per kilobase are true biallelic polymorphisms (i.e., A&E) due to sequence differences between the original parents (Su et al.

Table 1. Features of *T. gondii* chromosome Ia and Ib DNA sequence for the RH strain

	chromosome Ia	chromosome Ib
Total sequence length (bp)	1,923,819	2,013,089
% GC of genome	52.63	52.30
% GC of CDS	57.97	57.75
Number of predicted CDSs[a]	211	252
Number of pseudogenes	5	4
Expression confirmed by ESTs[b]	102 (47%)	117 (46%)
Average CDS length (bp)	2682	2185
Number of spliced genes	152 (69%)	192 (75%)
Average exon length (bp)	463	386
Average intron length (bp)	519	528
Average number of exons per gene	5.79	5.65
Gene density	1 per 9117 bp	1 per 7988 bp
Coding percentage	56.7%	58.1%
Number of tRNA genes	3	10
Number of SNPs[c]	332	12,925

[a]Excluding pseudogenes.
[b]Alignment to one or more ESTs.
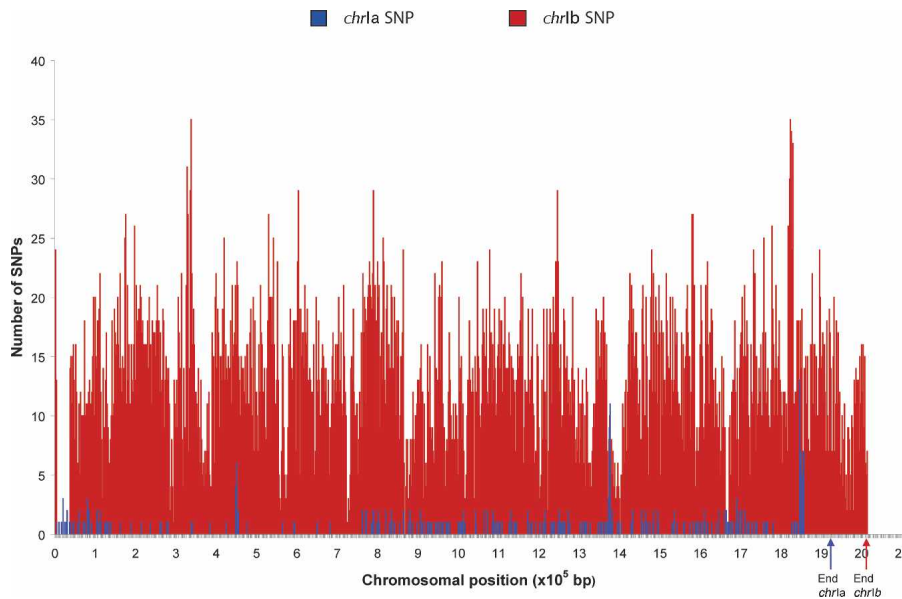[c]Based on a comparison between RH and ME49.

Figure 1. Single Nucleotide Polymorphisms (SNPs) on chromosome Ia vs. chromosome Ib for the RH vs. ME49 comparison. Distribution of SNPs (counts per 2000 bp). *chr*Ia has 332 SNPs, *chr*Ib has 12,925 SNPs.

version of *chr*Ia without apparent crossover. In order to confirm the uniparental inheritance of *chr*Ia among the three clonal lineages, we sequenced 12 of the 26 regions from *chr*Ia in two additional strains from each clonal type. These strains were selected from human or animal isolates that are representative of the three clonal lineages that predominate in the USA and France (Table 2). No SNPs were found in the sequence from these additional strains (a total of 9646 bp for each strain; see Fig. 2). These data are consistent with a common inheritance of *chr*Ia for all members of the clonal lineages, which predominate in North America and Europe.

Time of origin of the common *chr*Ia is coincident with origin of the clonal lineages

The combined data show two SNPs in 128,146 total bp, where 86,814 bp represents sequence from the 12 regions available for all nine strains (Fig. 2; Table 2). This very low rate of polymorphism of 2/86,814 bp for *chr*Ia is strikingly similar to the overall rate of unique SNPs between the three lineages. We have previously shown that while A&E polymorphisms occur across the genome at about 1/100 bp, when these biallelic polymorphisms are removed, the remaining unique polymorphism rate is on the order of 2/90,000 bp (Su et al. 2003). Based on the assumption that each strain evolves independently and these polymorphisms represent random (Poisson) neutral mutations that have occurred since their common ancestry (Rich et al. 1998), we have previously estimated that the three lineages originated ~10,000 yr ago (Su et al. 2003). Using this same model, the simplest explanation of our observations is that *chr*Ia was inherited in common by all three lineages at a time coincident with their recent origin.

Repression of meiotic recombination does not explain the monomorphic *chr*Ia

How might a virtually monomorphic *chr*Ia be established and maintained within an otherwise polymorphic genome? A monomorphic *chr*Ia is currently maintained in the population because the frequency of a sexual encounter with anything other than a member of a clonal lineage is very low (i.e., when crosses do occur, *chr*Ia is effectively "selfing" in the vast majority of crosses). However, this does not explain how the current monomorphic *chr*Ia was generated and initially propagated in a sexually reproducing polymorphic population. In sexually reproducing populations, population genetic theory suggests that chromosomes/chromosomal regions where meiotic recombination is suppressed may have reduced levels of polymorphism (Maynard-Smith and Haigh 1974; Charlesworth et al. 1993). This could be due to a selective sweep or purifying selection as proposed for the small fourth chromosome of *Drosophila melanogaster* (Berry et al. 1991; Jensen et al. 2002; Bartolome and Maside 2004), and similarly, it is possible that suppression of recombination could have led to the establishment of a monomorphic *chr*Ia in *T. gondii*. To
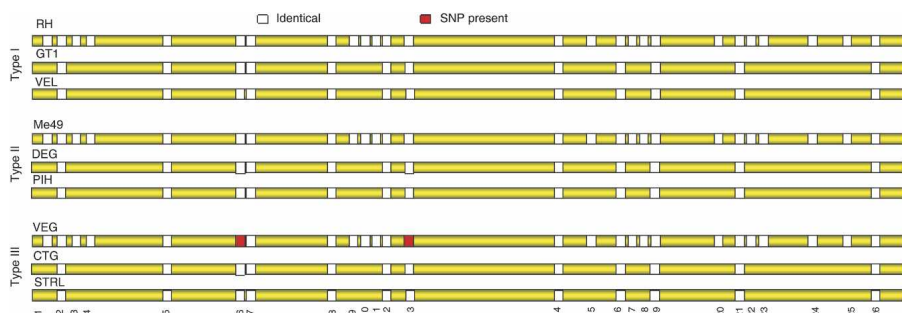
2003). Since this estimate may be biased because it is based solely on intron sequence, 53.6 kb of genomic sequence from a chromosome XII-derived BAC (RH) was generated. The alignment between RH and ME49 for this region estimates 6.60 SNPs per kilobase and predicts 13,289 SNPs due to the A&E dimorphism for *chr*Ib, again a figure consistent with the observed number of SNPs on *chr*Ib between these two strains. This frequency of A&E SNPs on *chr*Ib also agrees closely with that observed for ~200 genetic markers distributed across the genome (Su et al. 2002; Khan et al. 2005) and a more recent global analysis of strain-specific SNPs (Boyle et al. 2006). In stark contrast, *chr*Ia is virtually identical between RH and ME49, having only 0.17 SNPs per kilobase. This is likely to be an overestimate, because more than half of the apparent polymorphism is likely to be due to sequencing errors, even with the low sequencing error rate of ~1 in 10^4 bp achieved in this study. The 332 SNPs on *chr*Ia are not distributed at random across the chromosome (P~0), largely due to the paucity of SNPs between positions 150,000 and 750,000 and at the right-hand end of *chr*Ia (Fig. 1). The genes in the relatively polymorphism-free region in the center of the chromosome can be identified from the start positions given in Supplemental Table 3. There is no obvious bias in the functional categories of genes in these regions that would explain the lower polymorphism, although many of the genes are annotated as "unknown." The overall low rate of polymorphism argues that *chr*Ia is derived from a single ancestral chromosome that is extremely recent in origin.

*chr*Ia is shared in all lineages

In order to assess the generality of these observations, we analyzed a Type III strain called VEG, which represents the third predominant lineage. Since whole-genome sequence is not yet available, the EST database (Li 2003) was screened and 26 scattered regions of 800–900 bp that included at least three ESTs were selected for sequencing. There were only two SNPs in 23,654 bp of *chr*Ia sequenced from VEG (Fig. 2). This low rate of polymorphism argues that the Type III lineage also inherited the same

Figure 2. Composition of *chr*Ia for clonal (I, II, III) strains based on sequence analysis of 26 regions. Yellow represents areas not sequenced in the alternative strains, white represents sequenced areas that are identical to RH and ME49, and red represents regions containing a single SNP. Twelve regions are represented in all nine strains. Only two SNPs are seen in these 12 regions, which represents a total of 86,814 bp of sequence across the nine strains.

test whether a similar process might occur in *T. gondii*, we analyzed the segregation of SNPs on *chr*Ia in the progeny of a genetic cross between the Type I and Type III lineages as previously reported (Khan et al. 2005). Our analysis showed that 12 of the 31 progeny showed one or more crossing-over events in a single meiosis on *chr*Ia (Fig. 3) and, in experimental crosses, it is observed to have the highest frequency of recombination of any *T. gondii* chromosome (Khan et al. 2005). Thus, there is no physical suppression of recombination on *chr*Ia. Nevertheless, in experimental crosses of *T. gondii,* individual chromosomes can be inherited uniparentally (i.e., without crossing over). This raises the possibility that it is simply the result of chance that *chr*Ia is the only chromosome/chromosomal region common between the three clonal lineages, particularly if they are related by only a few crosses.

Chance inheritance of *chr*Ia is highly unlikely

A study of EST-identified SNPs by Boyle et al. (2006) revealed a surprising pattern of SNP distribution across the genome— haplotype blocks are conserved across large regions of the genome. Further analysis of these patterns indicates that the relationship between the clonal lineages might be explained by as few as two separate crosses. They propose that the existence of three clonal lineages were generated by two crosses, in which one lineage (Type II) is a common parent of the other two, and that the subsequent lineages have undergone little mixing since this recent origin. Given the two-cross model, it is expected that some chromosomes will not recombine and will be shared in their

entirety between the three lineages, some chromosomes will have shared regions, and some will have no shared regions. Experimental crosses with *T. gondii* predict a recombination rate of 0.4 (Khan et al. 2005); hence, the expected frequencies of these three classes are 0.09, 0.36, and 0.55, respectively. The probability of observing a single intact chromosome with no other chromosome having shared regions can be calculated using the trinomial distribution:

$$P = \frac{N!}{n_1!n_2!n_3!} p_1^{n_1} p_2^{n_2} p_3^{n_3}$$

where N is the number of chromosomes, n_1, n_2, n_3 are the observed numbers of shared chromosomes, chromosomes with shared sections, and chromosomes with no shared sections, and p_1, p_2, p_3 are the probabilities of observing these three types of chromosomes on the two-cross model.

Therefore:

$$P = \frac{14!}{1!0!n13!} 0.09^1 0.36^0 0.55^{13}$$

$$= 0.0005$$

Collapsing the first two classes to give a minimum expected number greater than five, the expected numbers of chromosomes with shared material and with no shared material are 7.7 and 6.3 compared with observations of 1 and 13, giving $\chi_1^2 = 8.1$ ($P = 0.004$). Given the two-cross model, the observed pattern is highly unlikely, with fewer shared chromosome/chromosomal regions than expected by chance. If more than two crosses led to the extant clonal lineages, the probability is further reduced, so it is extremely unlikely that this pattern would occur by chance.

Discussion

Comparative genome analysis of the two smallest chromosomes of *T. gondii*, *chr*Ia and *chr*Ib reveals a similar chromosome organization and content of genes and repeats between two of the three common lineages of *T. gondii*. Strikingly, analysis of the frequency of polymorphisms within these chromosomes reveals a markedly different pattern. *chr*Ib contains the expected frequency of SNPs seen in other regions of the genome of approximately one change per 100 bp. In contrast, the extremely low polymorphism rate on *chr*Ia argues that the Type I and Type II lineages acquired this entire chromosome from a common source. While our sampling of the Type III lineages and other representative strains is more limited, given the relatively low rate of recombination in *T. gondii*, it is unlikely that the intervening regions differ from this pattern. Hence, our data suggest that all members of the three clonal lineages that predominate in North America and Europe possess the identical *chr*Ia. Moreover, the estimated time of the establishment of this monomorphic chromosome is coincident with the origin of these clonal lineages ~10,000 yr ago (Su et al. 2003). Although the clonal lineages may be related by as few as two crosses, probability esti-

Table 2. Summary of clonal strains used for sequencing regions of chromosome Ia

Strain Name	ATCC #	Type	Date of isolation	Source	Host
RH	50838	I	1988[a]	USA	Human
GT1	50853	I	1980	USA	Goat
VEL	50852	I	1988	USA	Human
ME49 (B7 clone)	50840	II	1985[b]	USA	Sheep
DEG	50855	II	1987	France	Human
PIH	50857	II	1992	USA	Human
VEG	50861	III	1988	USA	Human
CTG	50842	III	1976	USA	Cat
STRL	50955	III	1993	USA	Human

[a]Cloned from original isolate (1939).
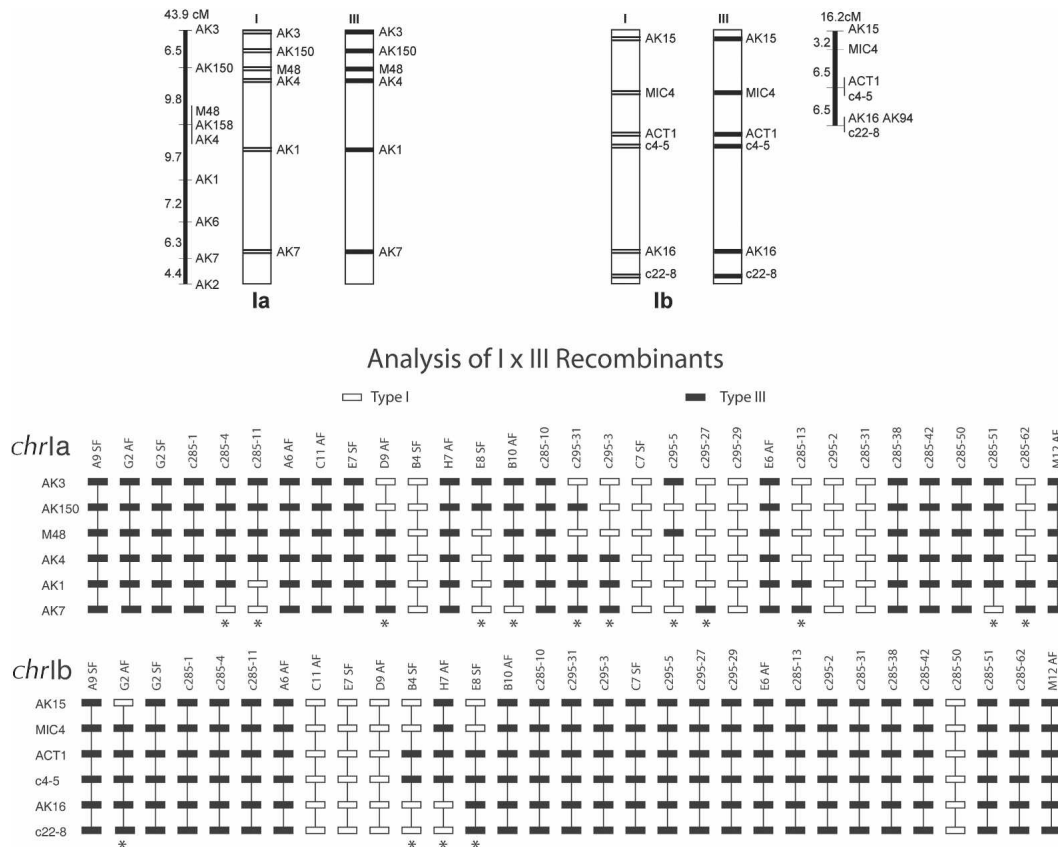[b]Cloned from original isolate (1965).

Figure 3. Progeny from genetic crosses between the Type I and Type III lineage were analyzed using six polymorphic markers on each of *chr*Ia and *chr*Ib. The resulting allele patterns are shown for each of 31 progeny at each locus (solid boxes, Type III lineage; open boxes, Type I lineage). Progeny where recombination has occurred are indicated at the *bottom* by an asterisk. The maps at the *top* represent the linear genetic maps for each chromosome with the markers shown on the *right* and the genetic distances shown on the *left* in centiMorgans.

mates indicate that the common inheritance of *chr*Ia alone is extremely unlikely to have arisen by chance. Rather, our analyses support a model where natural selection acting on the products of a recombination event in the wild led to the unusual fixation of *chr*Ia.

Despite the results from the experimental cross that argue that *chr*Ia monomorphism is not due to chromosome-specific repression of meiotic recombination, variations of a positive selective sweep model remain a possibility. Examination of a cross between relatively avirulent Type II and Type III *T. gondii* strains demonstrates that recombination can generate progeny far more virulent than either parent (Grigg et al. 2001a; Grigg and Suzuki 2003); not surprisingly, virulence involves multiple genes, as shown by linkage mapping studies (Su et al. 2002). The monomorphic *chr*Ia could be the result of just such a recombination process, in which the generation of a particular combination of alleles from different parents led to successful expansion of the clonal lineages bearing this unique chromosome. While the precise mechanism at play is uncertain, one plausible model suggests that this event first allowed or greatly enhanced asexual oral transmission between intermediate hosts (Su et al. 2003). Parasites capable of such asexual oral transmission would be expected to expand rapidly via clonal propagation, thus limiting, but not excluding exposure to sexual, meiotic recombination. The special combination of alleles on *chr*Ia would subsequently be maintained together by asexual growth, while still allowing the

rare opportunity for recombination to establish related but distinct lineages (Fig. 4). At present, we cannot predict which gene or genes on *chr*Ia might be responsible for the success of the clonal lineages, due to the fact that this chromosome has been maintained in entirety in all of the strains we have examined (all regions are equally associated with the phenotype; annotated genes are listed in Supplemental Table 3). Successful testing of this model will rely on identifying ancestral or exotic strains that express a wider range of asexual oral transmission phenotypes that we predict will have more variable *chr*Ia genotypes.

Asexual transmission now affords *T. gondii* with population dynamics more akin to zoonotic viruses, where rare recombinants can spread via direct horizontal transmission (Worobey and Holmes 1999). This has reached a culmination in the example of the three clonal lineages that predominate in North America and Europe, where widely different geographic regions share identical mutilocus genotypes. However, recent studies have revealed that the population structure of *T. gondii* in South America differs from this pattern and appears genetically more divergent (Ajzenberg et al. 2004; Lehmann et al. 2004; Khan et al. 2006; Su et al. 2006). It is presently uncertain whether a clonal population structure also predominates in these regions, albeit dominated by different genotypes, or whether they are truly panmictic. More severe forms of toxoplasmosis in humans have been associated with so-called atypical strains from such regions (Darde et al. 1998; Grigg et al. 2001b; Khan et al. 2006) and it is
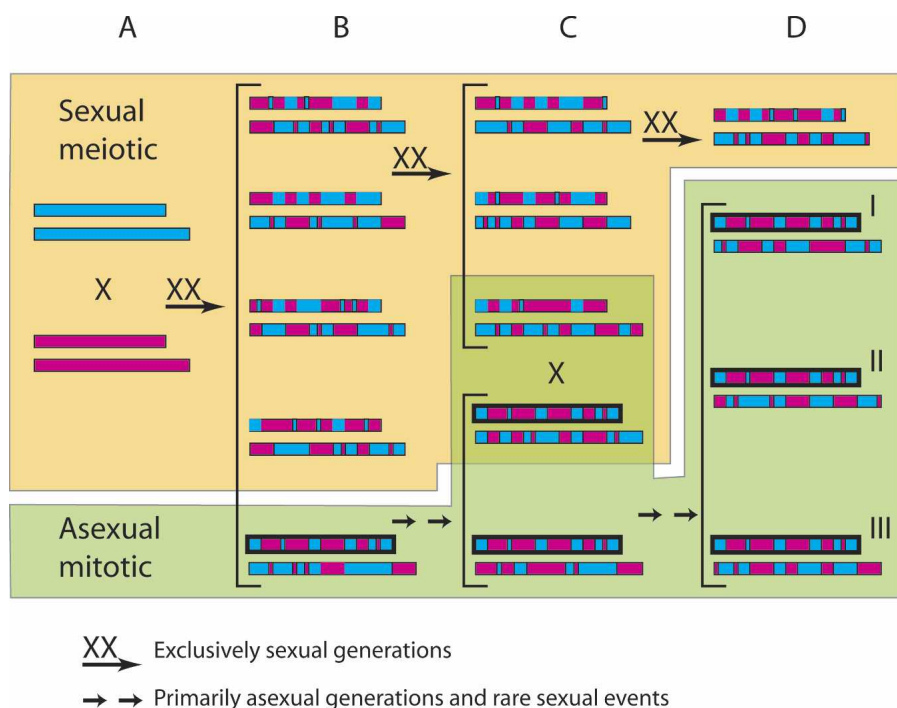
XX⟶ Exclusively sexual generations

→ → Primarily asexual generations and rare sexual events

Figure 4. Model of the evolutionary histories of *chr*Ia and *chr*Ib. (*A*) Original A&E cross. (A) blue; (E) pink; (shorter bar) *chr*Ia; (longer bar) *chr*Ib. (*B*) Recombinant progeny from a restricted gene pool. A special *chr*Ia (black outline), generated by recombination, carries allele(s) that allow or enhance transmission between intermediate hosts, permitting asexual propagation and suppressing recombination events that might break up the special combination of alleles. (*C*) Asexual growth expands the lineage in the population. Infrequent sexual events with highly recombinant relatives from the original cross allow parasites with the special chromosome Ia to acquire highly recombinant chromosomes including chromosome Ib. These new recombinant parasites can now expand in the population through asexual growth. (*D*) Natural selection for asexual growth maintains the special *chr*Ia intact and three different clonal lineages (i.e., Types I, II, and III) expand and become predominant in the population.

uncertain to what extent these strains share the monomorphic *chr*Ia described here. Despite a recently derived and highly clonal population structure, the continued recombination of *T. gondii* in the wild, while infrequent, may lead to strains that acquire new pathogenic mechanisms and can readily expand in the population. Further investigations into the genetic basis of this and other mechanisms that profoundly alter transmissibility and virulence will provide a greater understanding of the evolution of toxoplasmosis and global parasitic disease in general.

Methods

Parasite culture for DNA preparation

Tachyzoites of the RH (American Type Culture Collection [ATCC] # 50174) strain were propagated in monolayers of human foreskin fibroblasts as described previously (Su et al. 2003). Parasite harvesting, DNA isolation, and shotgun library preparation for *chr*Ia and *chr*Ib are described in the Supplemental material.

Whole-chromosome sequencing and annotation

Whole-chromosome shotgun sequencing (Hall et al. 2003) was used to determine the 1.92 and 2.01 Mb DNA sequences of *chr*Ia and *chr*Ib for *T. gondii* strain RH. The methods used in annotating the RH sequence are provided in the Supplemental material.

Comparison of strain RH with strain ME49

The manually finished sequence of *chr*Ia and *chr*Ib from the RH strain (http://www.sanger.ac.uk/Projects/ T_gondii/) was compared with a whole-genome shotgun assembly of the Type II ME49 (B7 clone) strain generated by TIGR (preliminary sequence data was obtained from The Institute for Genomic Research Web site at http://www. tigr.org). The polymorphisms were identified and classified as SNPs, insertion/ deletions, and replacements (Supplemental material). Su et al. (2003) analyzed the frequency of SNPs in noncoding regions (including 11 introns and the ITS1 region) in four Type I, three Type II, and three Type III strains. The frequency of SNPs between Type I and Type II strains was estimated from Table S3 of their Supplemental online materials as 28 SNPs per 4067 bp or 6.9 SNPs per kilobase compared.

The SNPs are expected to follow a Poisson distribution if they are randomly distributed across *chr*Ia. The chromosome was divided into "bins" and the number of SNPs in each bin were counted. Agreement with the Poisson distribution was assessed using the Poisson dispersion test. The result was not sensitive to the bin size chosen.

Recombination parameters for *chr*Ia and *chr*Ib

To identify genetic markers for *chr*Ia, we analyzed 10-kb sequences spaced ~500 kb apart along the chromosome. We compared the genomes of Type I and Type II strains using CLUSTALX (Higgins et al. 1996) to identify potential SNPs. Once SNPs had been identified, we amplified and sequenced genomic regions flanking SNPs from all three lineages using the type strains—RH (Type I), ME49 (Type II), and VEG (Type III) (Supplemental material). CLUSTALX was used to align the sequences and identify strain-specific polymorphisms. Further analysis of these markers is described in the Supplemental material. In total, nine polymorphic markers were developed on *chr*Ia. However, due to the biallelic pattern of *T. gondii*, only six of these detect differences between the Type I and Type III lineages, while three of them detect differences between the Type II and Type III lineages. Due to the larger number of markers and progeny available for the IxIII cross, we analyzed the segregation of markers in the 31 progeny from a previous cross (Su et al. 2002). Progeny were analyzed for segregation of polymorphisms on *chr*Ia and *chr*Ib by PCR amplification, restriction enzyme digestion, and gel electrophoresis to score alleles. The resulting genotypes were analyzed using MapMAKER (Lander et al. 1987) to generate genetic linkage maps of the chromosomes.

Acknowledgments

References

Ajioka, J.W., Boothroyd, J.C., Brunk, B.P., Hehl, A., Hillier, L., Manger, I.D., Marra, M., Overton, G.C., Roos, D.S., Wan, K.L., et al. 1998. Gene discovery by EST sequencing in *Toxoplasma gondii* reveals sequences restricted to the Apicomplexa. *Genome Res.* **8:** 18–28.

Ajzenberg, D., Banuls, A.L., Su, C., Dumetre, A., Demar, M., Carme, B., and Darde, M.L. 2004. Genetic diversity, clonality and sexuality in *Toxoplasma gondii*. *Int. J. Parasitol.* **34:** 1185–1196.

Bartolome, C. and Maside, X. 2004. The lack of recombination drives the fixation of transposable elements on the fourth chromosome of *Drosophila melanogaster*. *Genet. Res.* **83:** 91–100.

Berry, A.J., Ajioka, J.W., and Kreitman, M. 1991. Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* **129:** 1111–1117.

Boyle, J.P., Rajasekar, B., Saeij, J.P., Ajioka, J.W., Berriman, M., Paulsen, I., Roos, D.S., Sibley, L.D., White, M.W., and Boothroyd, J.C. 2006. Just one cross appears capable of dramatically altering the population biology of an eukaryotic pathogen like *Toxoplasma gondii*. *Proc. Natl. Acad. Sci.* **103:** 10514–10519.

Charlesworth, B., Morgan, M.T., and Charlesworth, D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134:** 1289–1303.

Darde, M.L., Bouteille, B., and Pestre-Alexandre, M. 1992. Isoenzyme analysis of 35 *Toxoplasma gondii* isolates and the biological and epidemiological implications. *J. Parasitol.* **78:** 786–794.

Darde, M.L., Villena, I., Pinon, J.M., and Beguinot, I. 1998. Severe toxoplasmosis caused by a *Toxoplasma gondii* strain with a new isoenzyme type acquired in French Guyana. *J. Clin. Microbiol.* **36:** 324.

Dubey, J.P. and Beattie, C.P. 1988. *Toxoplasmosis of animals and man*. CRC Press, Boca Raton, FL.

Grigg, M.E. and Suzuki, Y. 2003. Sexual recombination and clonal evolution of virulence in *Toxoplasma*. *Microbes Infect.* **5:** 685–690.

Grigg, M.E., Bonnefoy, S., Hehl, A.B., Suzuki, Y., and Boothroyd, J.C. 2001a. Success and virulence in *Toxoplasma* as the result of sexual recombination between two distinct ancestries. *Science* **294:** 161–165.

Grigg, M.E., Ganatra, J., Boothroyd, J.C., and Margolis, T.P. 2001b. Unusual abundance of atypical strains associated with human ocular toxoplasmosis. *J. Infect. Dis.* **184:** 633–639.

Hall, S.M., Ryan, M., and Buxton, D. 2001. Epidemiology. In *Toxoplasmosis: A comprehensive clinical guide* (eds. D.H.M. Joynson and T.G. Wreghitt), pp. 58–124. Cambridge University Press, Cambridge, UK.

Hall, N., Berriman, M., Lennard, N.J., Harris, B.R., Hertz-Fowler, C., Bart-Delabesse, E.N., Gerrard, C.S., Atkin, R.J., Barron, A.J., Bowman, S., et al. 2003. The DNA sequence of chromosome I of an African trypanosome: Gene content, chromosome organisation, recombination and polymorphism. *Nucleic Acids Res.* **31:** 4864–4873.

Higgins, D.G., Thompson, J.D., and Gibson, T.J. 1996. Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.* **266:** 383–402.

Howe, D.K. and Sibley, L.D. 1995. *Toxoplasma gondii* comprises three clonal lineages: Correlation of parasite genotype with human disease. *J. Infect. Dis.* **172:** 1561–1566.

Jensen, M.A., Charlesworth, B., and Kreitman, M. 2002. Patterns of genetic variation at a chromosome 4 locus of *Drosophila melanogaster* and *D. simulans*. *Genetics* **160:** 493–507.

Khan, A., Taylor, S., Su, C., Mackey, A.J., Boyle, J., Cole, R., Glover, D., Tang, K., Paulsen, I.T., Berriman, M., et al. 2005. Composite genome map and recombination parameters derived from three archetypal lineages of *Toxoplasma gondii*. *Nucleic Acids Res.* **33:** 2980–2992.

Khan, A., Jordan, C., Muccioli, C., Vallochi, A.L., Rizzo, L.V., Belfort Jr., R., Vitor, R.W., Silveira, C., and Sibley, L.D. 2006. Genetic divergence of *Toxoplasma gondii* strains associated with ocular toxoplasmosis, Brazil. *Emerg. Infect. Dis.* **12:** 942–949.

Kyes, S., Horrocks, P., and Newbold, C. 2001. Antigenic variation at the infected red cell surface in malaria. *Annu. Rev. Microbiol.* **55:** 673–707.

Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S.E., and Newburg, L. 1987. MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1:** 174–181.

Lehmann, T., Graham, D.H., Dahl, E.R., Bahia-Oliveira, L.M.G., Gennari, S.M., and Dubey, J.P. 2004. Variation in the structure of *Toxoplasma gondii* and the roles of selfing, drift, and epistatic selection in maintaining linkage disequilibria. *Infect. Genet. Evol.* **4:** 107–114.

Li, L., Brunk, B.P., Kissinger, J.C., Pape, D., Tang, K., Cole, R.H., Martin, J., Wylie, T., Dante, M., Fogarty, S.J., et al. 2003. Gene discovery in the apicomplexa as revealed by EST sequencing and assembly of a comparative gene database. *Genome Res.* **13:** 443–454.

Mackinnon, M.J. and Read, A.F. 2004. Virulence in malaria: An evolutionary viewpoint. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359:** 965–986.

Maynard-Smith, J. 1978. *The Evolution of Sex*. Cambridge University Press, Cambridge, UK.

Maynard-Smith, J. and Haigh, J. 1974. The hitchhiking effect of a favorable gene. *Genet. Res.* **23:** 23–35.

Rich, S.M., Licht, M.C., Hudson, R.R., and Ayala, F.J. 1998. Malaria's Eve: Evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci.* **95:** 4425–4430.

Sibley, L.D. and Boothroyd, J.C. 1992. Virulent strains of *Toxoplasma gondii* comprise a single clonal lineage. *Nature* **359:** 82–85.

Su, C., Howe, D.K., Dubey, J.P., Ajioka, J.W., and Sibley, L.D. 2002. Identification of quantitative trait loci controlling acute virulence in *Toxoplasma gondii*. *Proc. Natl. Acad. Sci.* **99:** 10753–10758.

Su, C., Evans, D., Cole, R.H., Kissinger, J.C., Ajioka, J.W., and Sibley, L.D. 2003. Recent expansion of *Toxoplasma* through enhanced oral transmission. *Science* **299:** 414–416.

Su, C., Zhang, X., and Dubey, J.P. 2006. Genotyping of *Toxoplasma gondii* by multilocus PCR-RFLP markers: A high resolution and simple method for identification of parasites. *Int. J. Parasitol.* **36:** 841–848.

Wan, K.L., Blackwell, J.M., and Ajioka, J.W. 1996. *Toxoplasma gondii* expressed sequence tags: Insight into tachyzoite gene expression. *Mol. Biochem. Parasitol.* **75:** 179–186.

Worobey, M. and Holmes, E.C. 1999. Evolutionary aspects of recombination in RNA viruses. *J. Gen. Virol.* **80:** 2535–2543.